

Algoritmos Supervisados

Discretize (Discretización)

Este algoritmo convierte las características numéricas en características categóricas, dividiendo el rango de valores continuos en intervalos discretos. Discretizar las características numéricas como purchase_lead, length_of_stay, flight_hour y flight_duration puede ser útil para ciertos algoritmos de machine learning que funcionan mejor con datos categóricos. Esto también puede ayudar a manejar valores atípicos y mejorar la interpretación de los modelos.



Resample (Remuestrear)

El algoritmo Resample se utiliza para modificar el tamaño del dataset, a menudo para equilibrar clases desbalanceadas. En nuestro dataset, donde puede haber una desigualdad en la cantidad de reservas completadas (booking_complete), este algoritmo puede ser útil para crear un conjunto de datos más equilibrado y permitir un entrenamiento más justo del modelo de machine learning.



Algoritmos no supervisados

Normalize (Normalizar)

Normalize se utiliza para escalar los valores de las características numéricas a un rango específico, típicamente entre 0 y 1. Esto es beneficioso para algoritmos sensibles a la escala de los datos, como las redes neuronales y los algoritmos basados en la distancia. En nuestro dataset, purchase_lead, length_of_stay, flight_hour y flight_duration son buenas candidatas para la normalización.



Principal Components (PCA)

Principal Components Analysis (PCA) se utiliza para la reducción de dimensionalidad, transformando un conjunto de variables posiblemente correlacionadas en un conjunto de valores de variables no correlacionadas llamadas componentes principales. Esto puede ayudar a reducir el ruido y mejorar la eficiencia computacional de los algoritmos de machine learning. En nuestro dataset, PCA puede ayudar a identificar las características más relevantes y reducir la complejidad.



